

## SOLICITAR UNA PÁGINA WEB

Una página web es un documento de texto plano, almacenado en el disco duro de una máquina que actúa como servidor, en el que se insertan pequeños fragmentos de código *html*, a los que llamamos etiquetas o *tags*. Las etiquetas *html* sirven para indicar al programa navegador qué formato debe dar al contenido de la página cuando la presente en la pantalla de la máquina cliente.

Para solicitar una página web, introducimos la dirección de la página o **URL** en la ventana de direcciones del navegador. El navegador pedirá en primer lugar, al servidor de nombres o **DNS**, que traduzca la URL a su **dirección IP**. Una vez en posesión de la dirección IP, cursará la petición directamente al servidor.



El servidor entregará al programa cliente el documento de texto, marcado con sus etiquetas, y todos los objetos, imágenes y archivos, citados en el mismo. El navegador o programa cliente reconstruirá en la pantalla del usuario la página con el formato especificado en las etiquetas y con los objetos en la posición y tamaño especificados.

Algunos proveedores de acceso a Internet mantienen servidores **proxy**, que entregan al navegador una copia de la página solicitada, evitando así que la petición llegue al servidor. Los servidores *proxy* son máquinas que almacenan copias de las páginas más solicitadas, interceptan las peticiones del cliente y le entregan la copia almacenada en su disco duro. Esto tiene algunas ventajas: ahorro de tráfico en la red, rapidez, filtrado de contenidos, etc. Pero también tiene algunos inconvenientes, entre los que cabe citar la falta de actualización o el filtrado de contenidos.

Es posible que, al solicitar una página web, el navegador no nos muestre la última versión de la página sino la versión almacenada en el disco del servidor *proxy*. Si deseamos que la petición llegue efectivamente al servidor, para que nos entregue la versión actualizada, hay que indicárselo al navegador pulsando el **botón Recargar** o las teclas **Ctrl + R**.

Desde la página web solicitada podemos visitar otras páginas haciendo clic en los **enlaces**, sin necesidad de escribir su dirección o URL en la ventana del navegador. Los enlaces aparecen destacados mediante subrayado y cambios de color en el texto de la página. Al hacer clic en un enlace se repite todo el proceso que nos mostrará la nueva página.

Podemos visitar de nuevo las páginas solicitadas durante una sesión. Para ir a una página que visitamos anteriormente pulsaremos en el botón **Atrás** o las teclas **Alt + ←**, para visitar de nuevo una página posterior pulsaremos el botón **Adelante** o las teclas **Alt + →**.

## CÓMO BUSCAR EN LA RED

El principal problema para un navegante, que se asoma a la Red para encontrar cierta información, es averiguar qué páginas o documentos contienen información pertinente para él. Este problema se ha resuelto de dos formas: utilizando la búsqueda indexada o un motor de búsqueda.

Los buscadores indexados o **directorios**, utilizan los datos relativos al título y descripción de una página web para incluirla en una base de datos indexada que, además, será verificada y actualizada por personas humanas. Estos directorios temáticos tienen la ventaja de que las búsquedas están fuertemente dirigidas hacia páginas que tratan del tema en el que estamos interesados. Entre sus inconvenientes hay que destacar que requieren más trabajo humano para su mantenimiento.

Los motores de búsqueda, **robots** de búsqueda o **spiders** son sistemas automáticos que buscan y analizan regularmente, cada cierto tiempo, en todas las páginas de la Red, acumulando en sus bases de datos información sobre su contenido. Cuando queremos localizar páginas con información sobre un tema, tecleamos un cierto número de palabras clave. El robot buscará en sus bases de datos y nos entregará una lista de páginas y documentos que contienen la palabra o palabras indicadas.

## A. BUSCADORES MÁS UTILIZADOS



Hay que citar, en primer lugar, a **Google** como el buscador más utilizado. **Google** fué el trabajo de doctorado de dos estudiantes de Informática de la Universidad de Stanford, Larry Page y Sergei Bin, que debió su éxito inicial al uso de algoritmos de búsqueda que clasifican las páginas, con un índice llamado **PageRank**, en función de su relevancia en la Red.

Actualmente, **Google** es un soporte publicitario formidable, una máquina de hacer dinero que ya no entrega resultados ajustados sólo a criterios de relevancia sino también a sus intereses comerciales. Tampoco entrega resultados independientes, tras haber aceptado aplicar filtros y mecanismos de censura en países como China, en los que los ciudadanos no tienen libertad de acceso a la Red, pero constituyen un mercado demasiado goloso.

Otro buscador poderoso es **Yahoo Search**, un buscador que utiliza un motor de búsqueda propio muy potente junto con un servicio de directorio.

Hay más buscadores interesantes: **Altavista**, que fué uno de los pioneros, sigue siendo un excelente buscador; **Alltheweb**; AskJeeves, que ahora se llama simplemente **Ask** e incluye una interesante barra de herramientas para hacer búsquedas especializadas.

También hay **metabuscadors**, es decir, motores de búsqueda que lanzan una petición a varios buscadores simultáneamente y presentan los resultados ordenados. Un metabuscador interesante es **Dogpile**, un robot que busca en **Google**, **Yahoo**, **MSN** y **Ask** simultáneamente. Otra apuesta interesante es la de **Kartoo**, un metabuscador que presenta los resultados gráficamente, en función de la relevancia de las páginas encontradas.

Finalmente, también se están desarrollando soluciones de código abierto en el terreno de los buscadores. El más relevante es **Nutch**, un buscador de software libre que puede instalarse sobre cualquier distribución GNU/Linux, que no está atado a una licencia propietaria.

## TÉCNICAS DE BÚSQUEDA EN LA WEB

Lo primero que hay que tener en cuenta es que el motor de búsqueda localiza, en su base de datos, **todas las páginas** que contienen la palabra clave. Si introducimos, por ejemplo, una palabra clave como **sopa**, el buscador localizará una enorme cantidad de páginas que contienen dicha palabra, entre las que pueden encontrarse expresiones como **sopa de letras**, **sopa boba**, **estar en la sopa**, **sopapo**, **hasta en la sopa**, etc.



Podemos precisar más la búsqueda, simplemente aumentando el número de palabras clave y escribiendo **sopa ajo**, **receta sopa ajo** o **receta sopa ajo castellana**. Aun así, el número de páginas localizadas, aunque menor es impresionante.



Los resultados de la búsqueda se presentan poniendo en primer lugar las páginas que incluyen en su contenidos **todas** las palabras clave y **en el mismo orden** en el que han sido escritas. Más adelante se mostrarán las páginas que incluyen dos o más de las palabras clave, empezando por las que aparecen en el mismo orden en que han sido escritas y, para terminar, en la lista aparecerán las páginas que incluyen una sola de las palabras clave. El buscador no entregará resultados redundantes si ya ha mostrado un número suficiente de ellos y nunca más de 1000 resultados.



Los algoritmos matemáticos de búsqueda son realmente complejos y valoran en torno a 150 criterios. Cuando busca las palabras clave en su base de datos, el robot **no tiene en cuenta las palabras superfluas** y elimina de los criterios de búsqueda las **partículas**, **pronombres** y **preposiciones** poco relevantes, las **mayúsculas**, los **acentos** y los **signos** de interrogación y exclamación. Si queremos que incluya en la búsqueda un término poco relevante, basta con poner delante el signo + como, por ejemplo, **quijote capítulo +I**.

También tiene en cuenta criterios de relevancia, para poner en primer lugar de la lista de resultados las páginas que contienen más información sobre las palabras solicitadas, las más visitadas y las que son enlazadas más frecuentemente desde otras páginas web. El algoritmo que calcula estos criterios de relevancia, el llamado **PageRank**, es un secreto comercial patentado por **Google**.

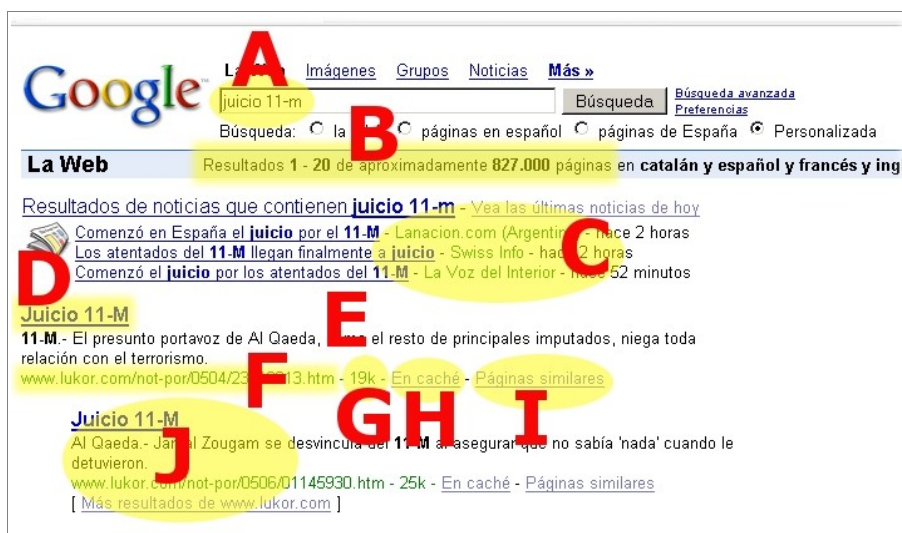
Cuando los términos de búsqueda se ponen entre comillas, como **"receta de sopa de ajo castellana"** por ejemplo, el buscador entregará exclusivamente los resultados que coinciden exactamente con la frase que se busca. Esta es pues una técnica que permite ajustar mucho la búsqueda y resulta útil para localizar un documento muy concreto como, **"orden 2265/2006, de 27 de diciembre, de la Consejería de Sanidad y Consumo"** por ejemplo. La búsqueda se ajusta tanto, con esta técnica, que es posible que no encuentre ningún resultado si el número de palabras entre comillas es excesivo, en cuyo caso conviene reducir el número de términos.



Conviene **evitar los acrónimos** cuando se ordena una búsqueda, porque el número de siglas coincidentes es, a veces, causa de errores. Es preferible teclear **organización mundial comercio**, en lugar de teclear **OMC**, simplemente porque hay otros organismos y empresas cuyas siglas pueden coincidir con las siglas tecleadas como palabra clave como, por ejemplo, la **Organización Médica Colegial** o la emisora **Onda Merlín Comunitaria**. Además, conviene utilizar el nombre no castellano de una organización, **World Trade Organisation** en nuestro ejemplo, si queremos obtener información relevante, procedente de la propia fuente y que no haya sido publicada en nuestro idioma.

## CÓMO INTERPRETAR LOS RESULTADOS DEL BUSCADOR

Los resultados de la búsqueda tienen información suficiente para saber si la página es relevante para nosotros:



### A. TÉRMINOS DE LA BÚSQUEDA

**Palabras clave** utilizadas para la búsqueda.

### B. ESTADÍSTICA

El buscador informa de la **cantidad** de resultados encontrados y del **tiempo** que ha necesitado.

### C. FUENTES DE INFORMACIÓN ESPECIALIZADAS

Cuando el buscador encuentra fuentes dedicadas, que parecen tener mucha información relevante sobre el tema propuesto, las presenta agrupadas al principio de la lista de resultados.

### D. TÍTULO DE LA PÁGINA

Cada resultado de búsqueda comienza con una línea con el **título** de la página. Esta línea **contiene un enlace** que conduce a nuestro navegador hasta ella. Si la página no tuviese título, porque no se haya completado en el campo <head>, aparecería su URL en la primera línea.

### E. FRAGMENTOS DE TEXTO

Son una o dos líneas de texto, en las que aparecen **resaltados** los términos de la búsqueda. Su lectura debe permitirnos valorar, **por el contexto** en el que aparecen las palabras clave, si el resultado se ajusta al tipo de documento que estamos buscando.

### F. URL DEL RESULTADO

Esta es la **dirección web** de la página encontrada.

### G. TAMAÑO DE LA PÁGINA

Tamaño del **archivo html**.

### H. VERSIÓN ALMACENADA EN CACHÉ

El robot de un buscador explora periódicamente la web y almacena, en una **caché**, una captura de pantalla de las páginas que han sido indexadas. Si por alguna razón, el **vínculo** de la primera línea no conduce al navegador hasta la página actual, se puede consultar la versión anterior almacenada en la **caché**, con una nota del día y la hora en la que se registró.

## I. PÁGINAS RELACIONADAS

Este enlace conduce a una lista de páginas que podrían tener alguna relación con la página del resultado, pero las técnicas actuales de análisis semántico e inteligencia artificial no producen resultados muy relevantes en este aspecto.

## J. OTROS RESULTADOS JERARQUIZADOS

Quando el buscador encuentra más resultados, almacenados *en el mismo sitio web*, los muestra a continuación, pero *escalonados* respecto al resultado más relevante.

## OTROS SERVICIOS DE BÚSQUEDA

Un buen buscador no sólo encuentra páginas web que contienen documentos relacionados con los términos de búsqueda. También es capaz de localizar imágenes, grupos de personas, noticias y muchas cosas más:

### A. DICCIONARIO

Una de la aplicaciones de un buscador, no demasiado conocida, es la posibilidad de utilizarlo como diccionario, para encontrar el significado de un término. Basta con teclear la palabra buscada, precedida de la orden **define:**. Así, por ejemplo, si tecleamos **define:acrónimo** obtenemos una recopilación de las definiciones de dicho término en varios diccionarios *on-line*:



### B. IMÁGENES

En la web hay almacenadas cantidades ingentes de imágenes. Un buscador puede localizarlas si se le suministran los criterios adecuados de búsqueda. Este es el resultado:



### A. TÉRMINO O TÉRMINOS DE LA BÚSQUEDA.



El buscador localizará imágenes que tengan, en su título o en su texto alternativo, los términos de la búsqueda. Cuantos más palabras clave utilicemos más afinados serán los resultados. Buscar con palabras clave en otros idiomas producirá también muchos más resultados.

## B. AJUSTE DE LAS PREFERENCIAS

El buscador de imágenes permite filtrar los resultados, entre otros ajustes. Cuando se trabaja con niños, es necesario activar los filtros para evitar que aparezcan en los resultados imágenes ofensivas.

## C. TAMAÑO DE LAS IMÁGENES

En los resultados hay imágenes de varios tamaños. Desde esta ventana puede seleccionarse el tamaño deseado para las imágenes buscadas: grandes, medianas y pequeñas.

## D. ESTADÍSTICA

Número de resultados obtenidos y tiempo invertido en la búsqueda.

## E. TEXTO ALTERNATIVO

Cuando el autor de la página ha establecido un texto alternativo para la imagen, aparecerá como pie de foto.

## F. NOMBRE DEL ARCHIVO

Si no hay texto alternativo, el buscador mostrará el nombre del archivo.

## G. INFORMACIÓN DE LA IMAGEN

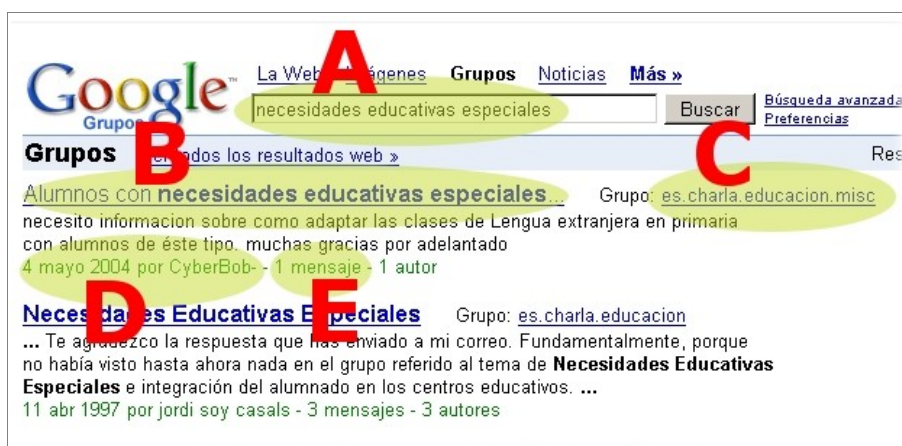
Posicionando el ratón sobre la imagen se muestra la información disponible: nombre y extensión del archivo, tamaño en bytes y píxeles, dirección de la página en la que aparece.

Al seleccionar uno de los resultados se abrirá una nueva página, en la que aparece una miniatura de la imagen con toda la información disponible y la página web de la que procede. Un enlace en la parte superior nos permite ver la imagen a tamaño completo y, si los permisos de copia lo permiten, copiarla en nuestro disco duro.

## C. BUSCAR GRUPOS AFINES

Tanto **Google** como **Yahoo** ofrecen servicios de software para grupos, que utilizan todavía el protocolo **Usenet** creado en 1979 para grupos de noticias. Con este servicio, un grupos de personas que tienen intereses comunes pueden mantener un foro de mensajes, intercambiar archivos, compartir fotos e imágenes y un largo etcétera.

Utilizando el servicio de búsqueda de grupos, el buscador puede localizar mensajes relacionados con los términos de búsqueda.

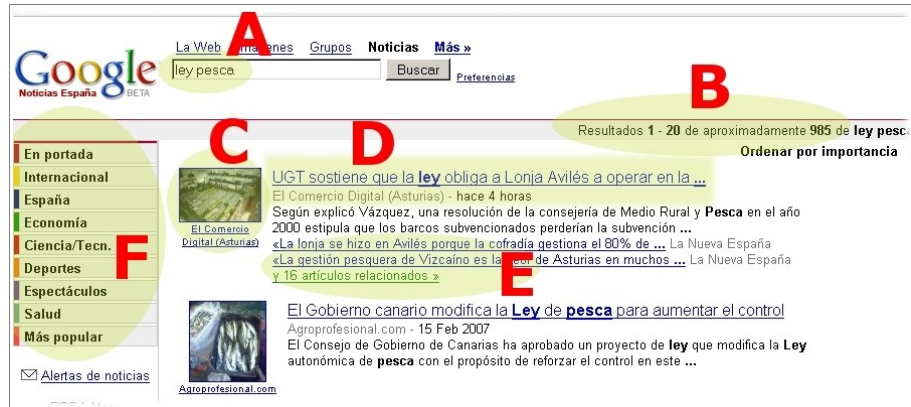


- A. Términos de la búsqueda.
- B. Asunto o hilo de conversación.
- C. Nombre del grupo.
- D. Fecha de la último mensaje en este hilo.

E. Cantidad de mensajes aportados.

#### D. BUSCAR NOTICIAS SOBRE UN TEMA

Cuando un tema de actualidad nos interesa mucho, y queremos localizar todas las noticias que se han publicado en la web sobre ese tema, podemos utilizar el servicio de búsqueda de noticias. Está disponible en todos los buscadores, aunque algunos funcionan mejor que otros:



A Palabras clave

B Estadística de resultados

C Imagen asociada a la noticia

D Datos de la noticia: Título de la noticia, página web en la que aparece y tiempo transcurrido

E Otras fuentes relevantes de la misma noticia

F Índice temático de noticias

#### E. BUSCAR UN LUGAR

Hay en Internet varios servicios de información geográfica (**GIS**) que permiten localizar lugares y servicios en un mapa, calcular rutas entre dos puntos, etc. Entre los mejores cabe citar los de [ViaMichelin](#), [Google Maps](#) y [El Callejero](#).

**ViaMichelin**, como empresa que presta servicios de navegación sobre **GPS**, posee una base de datos cartográficos y de servicios turísticos excelente. Al teclear, por ejemplo, "**Calle del Rosal 4, Oviedo, Asturias**", el sistema **GIS** de **ViaMichelin** nos ofrece un mapa de **Oviedo** en el que aparece marcada la localización solicitada:



Desde aquí podemos iniciar varias tareas como, por ejemplo, iniciar un itinerario, buscar hoteles y restaurantes, enviar las coordenadas de ese punto a nuestro GPS, etc.

**Google** también presta un excelente servicio de búsqueda geográfica, denominado **Google maps**, que permite localizar un lugar cualquiera en nuestro planeta. Si tecleamos, por ejemplo, **Calle del Rosal, Oviedo** el sistema de información geográfica de **Google** nos presenta un mapa de Oviedo, en el que aparece señalada la calle buscada:



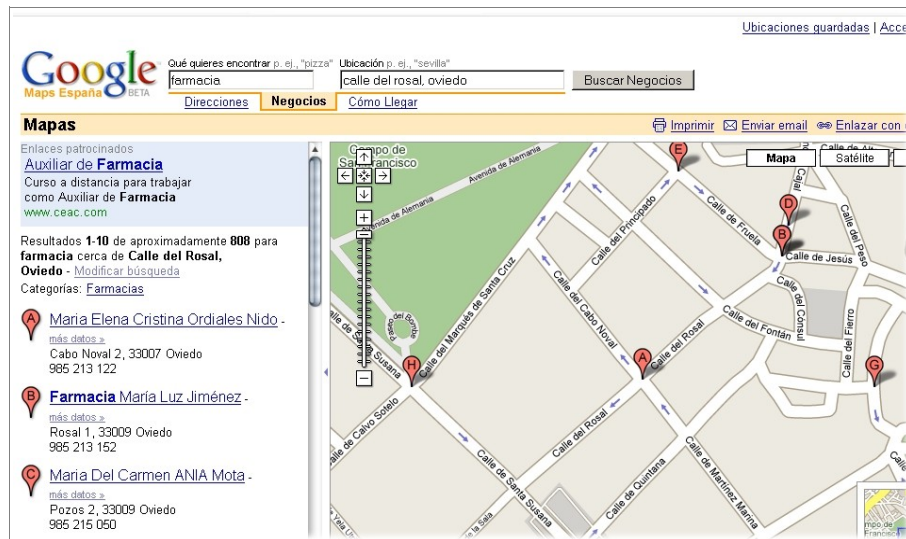
El usuario puede fijar este punto como el punto de partida predeterminado para sus rutas y sus búsquedas.

Los resultados de **Google maps** pueden ser visualizados de distintas formas, utilizando tres botones que aparecen en el extremo superior derecho de la imagen: como un **mapa convencional**, cuya escala puede variarse a voluntad utilizando el cursor que aparece a la izquierda; en **modo fotográfico**, para lo que **Google** utiliza una recopilación de fotografías hechas desde satélite y en modo mixto, superponiendo los nombres de las vías y calles principales sobre las imágenes de satélite.



Finalmente, **GoogleMaps** permite localizar empresas y servicios registrados y ubicarlos sobre el mapa. Por ejemplo, tecleando **Farmacias** sobre la zona que anteriormente hemos localizado en **Oviedo**, aparecen las farmacias de la zona señaladas con un icono rojo y, a la izquierda, un directorio con los datos de dichos establecimientos:





## F. TRADUCIR UNA PÁGINA WEB

A veces obtenemos, entre los resultados de una búsqueda, páginas en otros idiomas. **Google** ofrece un servicio de traducción automática, que está disponible si aparece en el resultado el enlace denominado *"Traduzca esta página"*:



El servicio está aún en fase de pruebas y los resultados no son muy buenos. El comportamiento de los traductores automáticos depende del desarrollo de las técnicas eficaces de inteligencia artificial y podemos esperar mejoras futuras.

## G. BUSCAR UN LIBRO Y LEERLO

Otro increíble proyecto de **Google** consiste en digitalizar libros, ya publicados en papel, para ponerlos al alcance del internauta. El servicio, denominado **GoogleBooks**, permite buscar en su base de datos para localizar un libro en concreto y, si está disponible, leerlo en pantalla con un programa lector o *e-reader*.

Por ejemplo, tecleando algunas palabras clave en el **buscador de libros de Google**, el buscador entregará una lista de libros que las contienen con algunos datos bibliográficos de cada resultado. También se pueden leer algunos fragmentos de texto.

Para obtener un acceso a algunas páginas completas del libro es necesario el permiso del autor o del editor y, si los derechos de autor del libro ya no están protegidos, se puede leer el texto completo.

Luis González  
Profesor de Tecnología